

REVIEW ARTICLE

Grading quality of evidence and strength of recommendations in clinical practice guidelines

Part 3 of 3. The GRADE approach to developing recommendations

J. L. Brożek¹, E. A. Akl^{1,2}, E. Compalati³, J. Kreis⁴, L. Terracciano⁵, A. Fiocchi⁵, E. Ueffing⁶, J. Andrews⁷, P. Alonso-Coello⁸, J. J. Meerpohl^{9,10}, D. M. Lang¹¹, R. Jaeschke¹, J. W. Williams Jr¹², B. Phillips¹³, A. Lethaby¹⁴, P. Bossuyt¹⁵, P. Glasziou¹⁶, M. Helfand¹⁷, J. Watine¹⁸, M. Afilalo¹⁹, V. Welch⁶, A. Montedori²⁰, I. Abraha²⁰, A. R. Horvath²¹, J. Bousquet^{22,23}, G. H. Guyatt¹ & H. J. Schünemann¹ for the GRADE Working Group

¹Departments of Clinical Epidemiology & Biostatistics and Medicine, McMaster University, Hamilton, ON, Canada; ²Departments of Medicine and Family Medicine, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY, USA; ³Allergy & Respiratory Diseases Clinic, Department of Internal Medicine, University of Genoa, Genoa, Italy; ⁴Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; ⁵Pediatric Division, Department of Child and Maternal Medicine, University of Milan Medical School at the Melloni Hospital, Milan, Italy; ⁶Centre for Global Health, Institute of Population Health, University of Ottawa, Ottawa, ON, Canada; ⁷Department of Obstetrics & Gynecology and Evidence-Based Practice Center, Vanderbilt University, Nashville, TN, USA; ⁸Iberoamerican Cochrane Centre, Clinical Epidemiology and Public Health Department, Institute of Biomedical Research-CIBER of Epidemiology and Public Health (CIBERESP-IIB Sant Pau), Barcelona, Spain; ⁹German Cochrane Centre, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany; ¹⁰Pediatric Hematology & Oncology, Center for Pediatrics and Adolescent Medicine, University Medical Center Freiburg, Freiburg, Germany; ¹¹Allergy/Immunology Section, Respiratory Institute, Cleveland Clinic, Cleveland, OH, USA; ¹²Department of Internal Medicine and Psychiatry, Duke University and Durham VA Medical Center, Durham, NC, USA; ¹³Centre for Reviews and Dissemination, University of York, York, UK; ¹⁴School of Population Health, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand; ¹⁵Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands; ¹⁶Centre for Evidence-Based Medicine, Department of Primary Health Care, University of Oxford, Oxford, UK; ¹⁷Portland Veterans Administration Medical Center and Department of Medicine, Oregon Health & Science University, Portland, OR, USA; ¹⁸Laboratoire de biologie polyvalente, Hôpital de Rodez, Rodez, France; ¹⁹Jewish General Hospital, McGill University, Montreal, QC, Canada; ²⁰Office of Regional Health Planning, Regional Health Authority of Umbria, Perugia, Italy; ²¹Screening and Test Evaluation Program, School of Public Health, University of Sydney and SEALS Department of Clinical Chemistry, Prince of Wales Hospital, Sydney, NSW, Australia; ²²Service des Maladies Respiratoires, Hôpital Arnaud de Villeneuve, Montpellier, France; ²³INSERM U1018, Villejuif, France

To cite this article: Brożek JL, Akl EA, Compalati E, Kreis J, Terracciano L, Fiocchi A, Ueffing E, Andrews J, Alonso-Coello P, Meerpohl JJ, Lang DM, Jaeschke R, Williams JW Jr, Phillips B, Lethaby A, Bossuyt P, Glasziou P, Helfand M, Watine J, Afilalo M, Welch V, Montedori A, Abraha I, Horvath AR, Bousquet J, Guyatt GH, Schünemann HJ, for the GRADE Working Group. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 3 of 3. The GRADE approach to developing recommendations. *Allergy* 2011; DOI:10.1111/j.1398-9995.2010.02530.x.

Keywords

clinical practice guidelines; evidence-based medicine; GRADE.

Correspondence

Holger J. Schünemann, MD, PhD,
Department of Clinical Epidemiology and
Biostatistics, Faculty of Health Sciences,
McMaster University, 1200 Main Street W,
Hamilton, ON, Canada.
Tel.: +1 (905) 525-9140 ext. 24931
Fax: +1 4433390565
E-mail: schuneh@mcmaster.ca

Accepted for publication 22 November 2010

DOI:10.1111/j.1398-9995.2010.02530.x

Edited by: Bodo Niggemann

Abstract

This is the third and last article in the series about the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to grading the quality of evidence and the strength of recommendations in clinical practice guidelines and its application in the field of allergy. We describe the factors that influence the strength of recommendations about the use of diagnostic, preventive and therapeutic interventions: the balance of desirable and undesirable consequences, the quality of a body of evidence related to a decision, patients' values and preferences, and considerations of resource use. We provide examples from two recently developed guidelines in the field of allergy that applied the GRADE approach. The main advantages of this approach are the focus on patient important outcomes, explicit consideration of patients' values and preferences, the systematic approach to collecting the evidence, the clear separation of the concepts of quality of evidence and strength of recommendations, and transparent reporting of the decision process. The focus on transparency facilitates understanding and implementation and should empower patients, clinicians and other health care professionals to make informed choices.

Abbreviations

ARIA, Allergic Rhinitis and its Impact on Asthma; DRACMA, Diagnosis and Rationale for Action against Cow's Milk Allergy; GRADE, Grades of Recommendation, Assessment, Development and Evaluation; SCIT, subcutaneous allergen-specific immunotherapy.

In the previous two articles in this series, we described the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to grading the quality of an available body of evidence supporting the choice of a therapeutic or diagnostic option for patients with allergic diseases (1, 2). In this article, we will describe how guideline panels apply the GRADE approach and use evidence to develop recommendations and how users of guidelines may interpret recommendations. Those making recommendations weigh benefits and downsides of the clinical management options being considered, integrating the information about the quality of a body of related evidence, patients' values and preferences as well as resource considerations.

What do clinicians want?

While it is important for clinicians to know the existing diagnostic or therapeutic approaches, they are primarily interested in, and often have questions about, which of the available options will achieve the most benefit and has the least downsides for their patients. They are also interested in how confident they can be about the balance of these desirable and undesirable consequences. Clinical practice guidelines should offer answers to these questions by advising about the most appropriate actions for 'typical' patients (there will always be, no matter how rare or frequent, situations in which a recommendation may not apply to an individual patient, because no recommendation can take into account all of the unique features of individual patients and circumstances). This advice – a recommendation – should be clear and unambiguous to be understood and applied correctly, and it should be developed transparently to be accepted (3–5).

Which option is better for patients?

The principle of beneficence (a moral obligation to act for the benefit of others, helping them to further their important and legitimate interests, often by preventing or removing possible harms) has been cherished by the medical community since antiquity, and it has been a mainstay of the literature of medical ethics in Western societies (6). It is also widely appreciated that beneficence in medical practice cannot be reduced to nonmaleficence (doing no harm). Those making recommendations must decide on the best course of action by balancing the desirable and undesirable consequences (i.e. benefits vs harms and burden) of possible management options. Because every decision in health care (as well as in life) comes with desirable and undesirable consequences, clinicians must be aware that the goal is to achieve the greatest benefit with minimum downsides. The desirable consequences of using a particular diagnostic or therapeutic approach include improved quality of life, reduced morbidity, longer survival and less resource use (costs). The undesirable consequences include complications from procedures and medications, consequences of incorrect diagnosis, adverse effects such as increased morbidity, anxiety, burden and higher resource use.

Take, for example, the question whether subcutaneous specific immunotherapy (SCIT) should be initiated in adults with allergic rhinitis, as opposed to management consisting only of allergen avoidance measures combined with regular medication use. The desirable consequences of SCIT might include reduced nasal and ocular symptoms on a long-term basis, reduced reliance on medications and improved quality of life. The undesirable consequences might include time spent and inconvenience associated with frequent visits to the doctor's office for subcutaneous injections, risk of anaphylactic reaction and higher initial resource use associated with immunotherapy.

Evaluating the balance between desirable and undesirable consequences of one action relative to another one (e.g. SCIT and usual symptomatic treatment vs usual symptomatic treatment alone) requires ascertaining the best estimates of the magnitude of the effects (e.g. by how much the symptoms will be reduced or the risk of systemic reactions from immunotherapy administration). It also requires judgments regarding the relative importance of these consequences for patients (e.g. how important the symptoms of rhinitis and the risk of asthma are to patients relative to the risk of systemic reactions, burden and cost; Fig. 1) (1).

Guideline panels judging the importance of outcomes on behalf of patients should ensure, to the best of their ability, that decisions reflect patients' values and preferences. They should make the reasoning underlying their judgments explicit and transparent (1, 5, 7). Indeed, explicit statements about the underlying values and preferences assumed by the guideline panel should be included when formulating recommendations (8). These statements should not be omitted when citing or translating recommendations. For instance, the Allergic Rhinitis and its Impact on Asthma (ARIA) guideline panel, following the GRADE approach, made a recommendation in favour of using SCIT in adults with allergic rhinitis. The

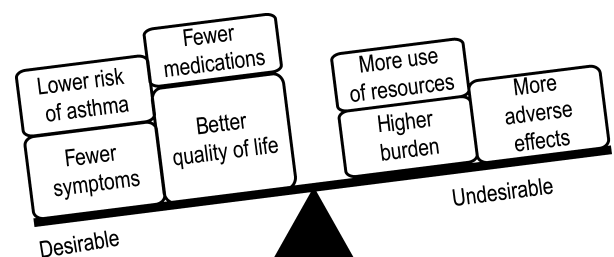


Figure 1 Balance of desirable and undesirable consequences of using subcutaneous allergen-specific immunotherapy (SCIT) with usual symptomatic treatment compared to usual symptomatic treatment alone in patients with allergic rhinitis. Those making recommendations have to consider the magnitude of the effects of SCIT and their relative importance to patients (depicted here as the size of each weight). The actual balance will depend on how a patient values particular outcomes, e.g. if a patient values avoiding adverse effects much more than depicted here, the balance will likely do be different and undesirable consequences may outweigh the desirable ones.

recommendation was complemented with the following statement: 'this recommendation placed a relatively high value on relieving the symptoms of allergic rhinitis, and a relatively low value on avoiding adverse effects and reducing resource expenditure' (9).

When the desirable consequences of following a management option (diagnostic, preventive or therapeutic) outweigh the undesirable ones, guideline panels would recommend that option. On the other hand, when the undesirable consequences outweigh the desirable ones, guideline panels would recommend against that management option (Fig. 2). To avoid making statements about what should not be done (e.g. 'we recommend that treatment A is not used'), they may recommend an alternative option stating what should be done (e.g. 'we recommend that treatment B is used rather than treatment A'). For example, the Diagnosis and Rationale for Action against Cow's Milk Allergy (DRACMA) guideline panel used the GRADE approach when evaluating the use of different infant formulas in children allergic to cow's milk proteins. The panel formulated a recommendation: 'in children with IgE-mediated cow's milk allergy, we suggest extensively hydrolysed milk formula rather than soy formula' (10).

Frequently the desirable and undesirable consequences of alternative management options are closely balanced, and the direction of a recommendation (i.e. for or against a given option) largely depends on assumed values and preferences of patients. Under these circumstances, guideline panels should make their assumptions about these preferences explicit. For instance, the ARIA guidelines recommended sublingual allergen-specific immunotherapy in adults with allergic rhinitis because of pollen. The authors stated that 'this recommendation placed a relatively high value on alleviating the symptoms of rhinitis, and relatively low value on avoiding adverse effects and resource expenditure' and the ARIA panel added a remark stating: 'local adverse effects are relatively frequent (approximately 35%); an alternative choice may be equally

reasonable, if patients' values or preferences differ from those described here' (9).

Is the chosen option truly better?

Those making recommendations may have higher or lower confidence that following their recommendation will have more desirable than undesirable consequences across the range of patients for whom the recommendation is intended (7). They inform users of guidelines (clinicians, patients and their family members, policy makers) about the degree of their confidence by specifying the strength of recommendations. While in reality, the balance between desirable and undesirable consequences is a continuum, the GRADE approach uses two grades of the strength of recommendations – strong or conditional – reflecting the confidence in the clarity of that balance or lack thereof. This dichotomy serves to simplify the message and improve understanding and communication as well as suggests clear implications for patients', clinicians' and policy makers' actions (Table 1). In various guidelines following the GRADE approach, words other than 'weak' have been used to express the lower confidence in the balance of desirable and undesirable consequences, e.g. 'qualified' or 'discretionary'.

Sometimes, authors of guidelines formulate their recommendations only as statements about the available evidence (e.g. chromones are effective in the treatment of allergic rhinitis), but do not explicitly specify what action should follow (e.g. should chromones be used in treatment of allergic rhinitis, given all other available treatment options?) (3). GRADE suggests phrasing recommendations in an active voice as clear indications what specific action should follow. For example, many guidelines developed following the GRADE approach worded their recommendations as 'we recommend...' and 'we suggest...' to distinguish strong from conditional recommendations. Alternative wording for strong

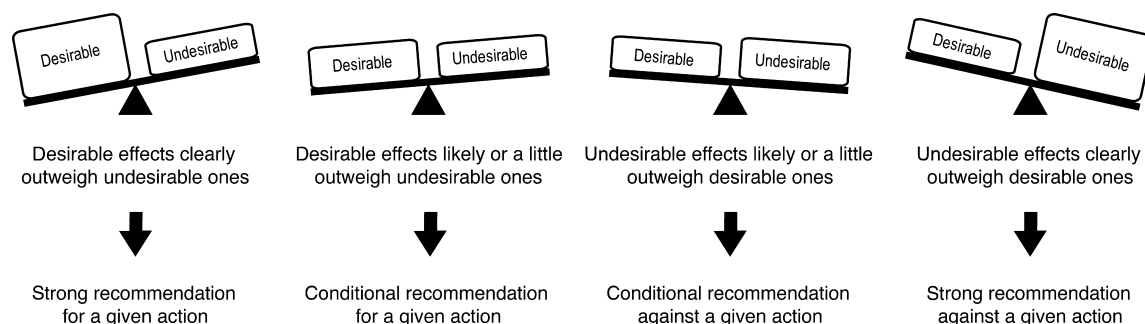


Figure 2 Balance of desirable and undesirable outcomes determines the direction and influences the strength of a recommendation. This balance depends on the magnitude of the expected desirable and undesirable effects and how patients value certain outcomes. If guideline panel is confident that desirable consequences outweigh the undesirable ones (leftmost situation), they are likely to make a strong recommendation for a given management option. On the contrary, if they are confident that undesirable

consequences outweigh the desirable ones (rightmost situation), they are likely to make a strong recommendation against it. However, if the guideline panel is not confident about the balance of desirable and undesirable consequences, they will make a weak recommendation. In such a case, the direction of recommendation (for or against certain intervention) depends on panel's judgement about which of the compared management options is likely to be better for the majority of patients.

Table 1 Implications of the two grades of strength of recommendations in the GRADE approach

Target group	Strong recommendations*	Conditional (weak) recommendations
Patients	Most people in your situation would want the recommended course of action and only a small proportion would not	The majority of people in your situation would want the recommended course of action, but many would not
Clinicians	Most patients should receive the recommended course of action	Recognize that different choices will be appropriate for different patients and that you must make greater effort to help each patient to arrive at a management decision consistent with his or her values and preferences; decision aids and shared decision making are particularly useful
Policy makers	The recommendation can be adopted as a policy in most situations	Policy making will require substantial debate and involvement of many stakeholders

GRADE, Grades of Recommendation, Assessment, Development and Evaluation.

*Strong recommendations based on high quality evidence will apply to most patients for whom these recommendations are made, but they may not apply to all patients in all conditions; no recommendation can take into account all of the unique features of individual patients and clinical circumstances.

recommendations includes ‘clinicians should...’ while weak recommendations can be phrased as ‘clinicians might...’ or ‘we conditionally recommend...’. The choice of words should reflect the intended strength of recommendation and its implications (Table 1). When developing, adapting, or translating recommendations, it is important to use words that are clearly understood and unambiguous in the given language and cultural milieu.

What determines the confidence that the chosen option is better?

In the GRADE approach, four factors influence the strength of recommendations: magnitude of the difference between the desirable and undesirable consequences, quality of the available supporting evidence, certainty about values and preferences of patients, and the resource expenditure associated with the compared management options.

Magnitude of the difference between the benefits and harms or burden

When the desirable consequences of following the recommendation clearly outweigh the undesirable consequences (Figs 1 and 2), it is more likely that the recommendation will be strong. When the desirable and undesirable consequences are closely balanced, a weak recommendation is warranted. While most original studies and systematic reviews present the magnitudes of effect of outcomes in relative terms (e.g., relative risk, hazard ratio, odds ratio), weighing the magnitude of the difference between the desirable and undesirable consequences to develop a recommendation also requires the knowledge of absolute effects for a specific population or situation. In certain situations, when values and preferences or attitude towards the resource use differ from those assumed by guideline developers, patients, clinicians and

policy makers may choose to examine the magnitude of effects of management options on the outcomes of interest themselves, rather than relying on judgements of those making the recommendation. For instance, in the example cited above, a guideline panel made a conditional recommendation to use extensively hydrolysed milk formula rather than soy formula in children with IgE-mediated cow’s milk allergy (10). The authors explicitly stated that they placed a relatively high value on avoiding adverse reactions to soy formula and a relatively low value on an inferior acceptance of the extensively hydrolysed milk formula and resource utilization. In settings where relative importance of resource expenditure is lower, one may choose to examine the supporting evidence (the magnitude of effects of extensively hydrolysed milk and soy formulas on symptoms of cow’s milk allergy, allergic reactions to formulas, nutritional status, acceptance, quality of life, cost, etc.) and decide on the balance of desirable and undesirable consequences oneself. The merit of this approach is in transparency and being explicit about the rationale for the choice.

Quality of the supporting evidence

The second factor that determines the strength of a recommendation is the quality of the body of supporting evidence for each of the critical outcomes (1, 2). Many other approaches confuse the strength of recommendations with the quality of evidence. Furthermore, in many approaches, rating of quality of evidence is usually limited to the consideration of the sole methodological design of the studies from which the evidence came (11). For example, some approaches would state that a recommendation is ‘level A’, which informs readers that a systematic review addressing that particular question is available. However, such a system of grading recommendations does not inform about the quality of that systematic review, the quality of the studies that were

included, the directness of the evidence (generalizability or applicability) or the confidence that patients will benefit from the treatment. To address this limitation, in the GRADE approach, the strength of a recommendation is separated clearly from the quality of evidence that, nonetheless, remains an important determinant of the strength of recommendations (1, 2, 7). If the quality of evidence is low, i.e. one is uncertain about the magnitude of particular desirable and undesirable consequences of an intervention, it is more likely that the recommendation will be weak.

Values and preferences

The third determinant of the strength of recommendations is variability in, or uncertainty about, preferences or values that patients assign to the outcomes of interest. Different patients can take different views about what outcome constitutes benefit or harm, and clinicians' understanding of importance of particular outcomes for patients can differ from that of the patients'. Explicit consideration of patients' values and preferences in making recommendations stems from acknowledgement of patients' liberty (autonomy). Alternative management strategies always have associated advantages and disadvantages, and thus, a trade-off is always necessary. How patients and guideline panel members value particular benefits, risks and inconvenience is critical to any recommendation and its strength. However, data about patients' preferences and values are often limited. It is ideal for clinicians to elicit the preferences and values directly from patients themselves. It would also be ideal for guideline panels to obtain estimates of values and preference from population-based studies, but such studies are often not available. To address this shortcoming, guideline panels often include patient representatives, seek input from selected patients or rely on panel members' judgement and their clinical expertise. Because values and preferences play a substantial role in determining the direction and strength of recommendations, GRADE urges guideline panels to state explicitly what values and preferences they considered and what weight they placed on each outcome. This transparent explanation facilitates the interpretation of recommendations, especially weak ones for which the best course of action is less certain. Table 2 provides examples of such statements.

On rare occasions, values that patients or guideline panel members assign to particular outcomes or their preferences are so strong that authors of the guidelines feel compelled to make a strong recommendation even in the face of low or very low quality evidence. For instance, the ARIA guideline panel made a strong recommendation against the use of intramuscular glucocorticosteroids for treatment of allergic rhinitis. Although this recommendation was supported by low-quality evidence, the ARIA panel members placed a very high value on avoiding potential rare serious adverse effects and a much lower value on their potential high efficacy and convenience of use (9).

On exceptional occasions, values or preferences for particular outcomes of those making recommendations are so

strong that they feel it is necessary to make a strong recommendation, even though they can rationally assume that patients' preferences may differ. For instance, the ARIA guideline panel made a strong recommendation that patients with asthma and/or allergic rhinitis who are allergic to animal dander avoid exposure to pet allergens at home (9). In this act of paternalism, the guideline panel intentionally recommended avoidance of animal dander despite suspected patient preferences (autonomy) justifying its decision by the goal of benefiting patients. An intense ethical debate has not yet reached consensus, whether beneficence could justify restricting patients' autonomy and, if so, on what basis (6). However, it seems that paternalism cannot be justified unless the risk of a preventable harm or loss of a benefit is significant, the action is likely to prevent the harm or obtain the benefit, anticipated benefits outweigh the risks and the least autonomy-restrictive action is proposed (6).

Resource use

Use of resources (cost) is another potentially important outcome associated with alternative management options, like mortality, morbidity and quality of life. A diagnostic or therapeutic intervention may increase resource use or decrease it compared to an alternative intervention. However, resource expenditure differs from other outcomes in several ways (12). Costs may vary widely among and even within countries, regions, or jurisdictions; they may have different implications in different settings and for patients with different socio-economical status and may quickly change over time. Therefore, it is prudent that guideline panels consider and document estimates of resource use, which could aid interpretation across settings and time.

Moreover, patients experience health improvements and bear the burden of adverse health outcomes, but health care costs and resource expenditure typically are shared by the society as a whole. Attitudes differ and the differences have not yet been resolved whether costs or resource expenditure should influence a clinician's decision about management of individual patients (social vs individual beneficence). Finally, matters relating to resource expenditure are highly political and may result in potential conflict of interest for some guideline panel members.

Attitudes differ as to whether resource use should influence a clinician's decision about treating individual patients. Depending on who bears the cost, it may be directly important for patients or important for the society as a whole. Therefore, although higher resource use will probably reduce the likelihood of a strong recommendation in favour of a particular intervention, the context of the recommendation will be critical. In considering resource allocation, those making recommendations must be very specific about the setting to which a recommendation applies and the perspective they took, i.e. that of a patient, a third party payer or society as a whole.

Table 2 presents examples of various recommendations from the two guidelines in the field of allergy that have been

Table 2 Examples of strong and conditional (weak) recommendations from the ARIA (9) and DRACMA (10) guidelines and the factors that primarily influenced the strength of recommendations

Factors influencing the strength of recommendations the most	Strong recommendations	Conditional (weak) recommendations
Balance of desirable and undesirable consequences	In children and pregnant women, we recommend total avoidance of environmental tobacco smoke (i.e. passive smoking) (strong recommendation I very low quality evidence). <i>Remarks</i> Smoking and exposure to second-hand smoke are common health problems around the world causing a substantial burden of disease for children and adults. While it is very rare to make a strong recommendation based on low or very low quality evidence, the ARIA guideline panel felt that in the absence of important undesirable effects of smoking cessation or reducing the exposure to second-hand smoke, the balance between the desirable and undesirable effects is clear. In settings where oral food challenge is considered a requirement for making a diagnosis of IgE-mediated cow's milk allergy (CMA), we recommend using oral food challenge with cow's milk as the only test without performing a skin prick test as a triage or an add-on test to establish a diagnosis (strong recommendation I very low quality evidence). <i>Underlying values and preferences</i> This recommendation places a relatively high value on avoiding resource consumption and the risk of anaphylactic reactions at home in patients who would be misclassified by a skin prick test alone. It places a lower value on anaphylactic reactions in a controlled setting that can be managed by experienced personnel when oral food challenge is performed. This recommendation also places a high value on avoiding any unnecessary treatment in patients who would be incorrectly classified by a skin prick test as allergic to cow's milk. <i>Remarks</i> This recommendation applies to clinical practice settings. In research settings there may be compelling reasons to perform skin prick tests even though a food challenge test with cow's milk is always being done.	We suggest new generation oral H1-antihistamines rather than intranasal H1-antihistamines in adults with seasonal allergic rhinitis (weak recommendation I moderate quality evidence). <i>Underlying values and preferences</i> This recommendation places a relatively high value on probable higher patient preference for oral vs intranasal route of administration as well as avoiding bitter taste of some intranasal H1-antihistamines, and relatively low value on increased somnolence with some new generation oral H1-antihistamines. In many patients with different values and preferences or those who experience adverse effects of new generation oral H1-antihistamines an alternative choice may be equally reasonable. In settings where oral food challenge is not considered a requirement in all patients suspected of IgE-mediated cow's milk allergy, in patients with high pretest probability of CMA we suggest using a skin prick test with a cut-off value of ≥ 3 mm as a triage test to avoid oral food challenge in those in whom the result of a skin prick test turns out positive (weak recommendation I low quality evidence). <i>Underlying values and preferences</i> This recommendation places a relatively high value on avoiding burden, resource use and very likely anaphylactic reactions during the oral food challenge test (approximately 50–70% food challenges avoided). It places a lower value on unnecessary treatment of around 1 in 20 patients misclassified as allergic to cow's milk (5–6% false positive results). <i>Remarks</i> A high pretest probability of CMA (approximately 80%) can be estimated based on the history and would represent, for instance, patients who experienced an anaphylactic reaction in the past.
Quality of evidence	In patients with allergic rhinitis, we recommend intranasal glucocorticosteroids rather than intranasal H1-antihistamines (strong recommendation I high quality evidence). <i>Underlying values and preferences</i> This recommendation places a relatively high value on efficacy of intranasal glucocorticosteroids, and a relatively low value on their rare adverse effects.	In patients allergic to indoor moulds, we suggest avoiding exposure to these allergens at home (weak recommendation I very low quality evidence). <i>Underlying values and preferences</i> This recommendation places a relatively high value on possible reduction in the symptoms of rhinitis and asthma, and a relatively low value on the burden and cost of interventions aimed at reducing exposure to household moulds.

Table 2 (Continued)

Factors influencing the strength of recommendations the most	Strong recommendations	Conditional (weak) recommendations
Values and preferences	In patients with allergic rhinitis because of animal dander, we recommend avoiding exposure to these allergens at home (strong recommendation very low quality evidence). <i>Underlying values and preferences</i> This recommendation places a relatively high value on potential reduction of symptoms of allergic rhinitis, and a relatively low value on psychosocial downsides of not having a pet or the inconvenience and cost of environmental control measures. <i>Remarks</i> Based on biological rationale, there is little doubt that total avoidance of animal allergens at home, and probably also marked reduction in their concentration, can improve symptoms, despite paucity of published data to substantiate this statement.	In patients with seasonal allergic rhinitis, we suggest intranasal glucocorticosteroids over oral H1-antihistamines in adults (weak recommendation low quality evidence). <i>Underlying values and preferences</i> This recommendation places a relatively high value on the likely higher efficacy of intranasal glucocorticosteroids. In many patients with strong preference for oral vs intranasal route of administration an alternative choice may be reasonable.

ARIA, Allergic Rhinitis and its Impact on Asthma; DRACMA, Diagnosis and Rationale for Action against Cow's Milk Allergy.

developed following the GRADE approach (9, 10) and the factors that influenced the strength of recommendations.

Recommendations to use interventions only in the context of research

Those making recommendations may face questions about the use of diagnostic or therapeutic interventions associated with potentially appreciable benefits, downsides or resource use but with insufficient evidence to support the informed decision. In such a case, one might think that any choice would not be justified and many panels choose not to make any recommendation. However, patients and clinicians still need advice about what is the better option. GRADE encourages guideline panels to make a recommendation despite paucity of evidence, because it is very likely that no one has spent as much time as the guideline panel on finding an answer. However, this paucity of evidence should be explicitly acknowledged.

On occasion, those making recommendations may have so little confidence in the effects of a certain intervention that they will refrain from making a recommendation. If the intervention is promising, they may recommend that further research is carried out. For instance, the DRACMA guideline panel considered the use of allergen microarrays in the diagnosis of cow's milk allergy. Because the technology may be useful but the guideline panel found very little evidence about its diagnostic performance, they recommended that allergen microarrays should be used only in the context of well-designed studies that investigate the accuracy of commercially available allergen microarrays compared to an oral food challenge test with cow's milk (10). Similarly, the same panel refrained from making a recommendation about the choice

between the use of soy formulae and extensively hydrolysed rice formulae in children with cow's milk allergy and called for more well-designed randomized trials comparing these milk substitutes.

Conclusions

In this series of three articles, we presented the GRADE approach to grading the quality of evidence and the strength of recommendations (1, 2). In this last article, we described how those following the GRADE approach integrate research evidence, patients' values and preferences, and consideration of resource use to transparently make sensible clinical recommendations. We discussed the factors that influence the strength of recommendations about the use of diagnostic, preventive and therapeutic interventions. We provided examples from two recently developed guidelines in the field of allergy that followed the GRADE approach.

The advantages of the GRADE approach are the focus on outcomes that are important to patients, explicit consideration of patients' values and preferences, the systematic approach to collecting the evidence, clear separation of the quality of evidence and strength of recommendations, and transparent reporting of the decision process.

The approach requires close collaboration of clinicians and methodologists with appropriate expertise. Ideally, the recommendations would be based on systematic reviews of available evidence. This requirement is shared by all approaches to developing clinical recommendations that strive to provide sensible, unbiased and unambiguous advice about the best clinical practice. The GRADE approach will likely evolve in parallel with the developments in clinical knowledge and research methodology.

The greatest merit of the GRADE approach to making recommendations is that it empowers patients, clinicians and other health care professionals to make informed choices. In addition, it facilitates understanding and implementation and emphasizes the importance of the input of all concerned parties: patients, caregivers, those with clinical expertise and those with expertise in the methods of clinical research and guideline development.

Acknowledgments

We thank Mr Jason Behrmann, MSc from the University of Montreal for his critical review of the manuscript. JLB holds a McMaster University Department of Medicine Internal Career Research Award. This work was partially funded by 'The human factor, mobility and Marie Curie Actions Scientist Reintegration' European Commission Grant: IGR 42192 – 'GRADE' to HJS.

References

1. Brożek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. *Allergy* 2009;**64**: 669–677.
2. Brożek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy* 2009;**64**:1109–1116.
3. Hussain T, Michel G, Shiffman RN. The Yale Guideline Recommendation Corpus: a representative sample of the knowledge content of guidelines. *Int J Med Inform* 2009; **78**:354–363.
4. Global Programme on Evidence for Health Policy. Guidelines for WHO guidelines [EIP/GPE/EQC/2003.1]. Geneva, Switzerland: World Health Organization, 2003.
5. Guyatt GH, Prasad K, Schünemann HJ, Jaeschke R, Cook DJ. How to use a patient management recommendation. In: Guyatt GH, Rennie D, Meade MO, Cook DJ, editors. *Users' guides to the medical literature: a manual for evidence-based clinical practice*, 2nd edn. McGraw-Hill Professional, 2008:597–618.
6. Beauchamp TL, Childress JF. In: *Principles of biomedical ethics*, 6th edn. New York: Oxford University Press, 2008.
7. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A et al. Going from evidence to recommendations. *BMJ* 2008;**336**:1049–1051.
8. Schünemann HJ, Munger H, Brower S, O'Donnell M, Crowther M, Cook D et al. Methodology for guideline development for the Seventh American College of Chest Physicians Conference on Antithrombotic and Thrombolytic Therapy: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest* 2004;**126**: 174S–178S.
9. Brożek JL, Bousquet J, Baena-Cagnani CE, Bonini S, Canonica GW, Casale TB et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines – 2010 Revision. *J Allergy Clin Immunol* 2010;**126**:466–476.
10. Fiocchi A, Brożek J, Schunemann HJ, Bahna SL, von Berg A, Beyer K et al. World Allergy Organization (WAO) Diagnosis and Rationale for Action against Cow's Milk Allergy (DRACMA) Guidelines. *World Allergy Organiz J* 2010;**3**:57–161.
11. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004;**4**:38.
12. Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A et al. Incorporating considerations of resources use into grading recommendations. *BMJ* 2008;**336**: 1170–1173.